# Circulating vitamin D concentrations and risk of breast and prostate cancer: a Mendelian randomization study

Xia Jiang,[1,2#] Niki L Dimou,[3#] Kawthar Al-Dabhani,[4] Sarah J Lewis,[5,6] Richard M Martin,[5-7] Philip C Haycock,[5,6] Marc J Gunter,[8] Timothy J Key,[9] Rosalind A Eeles,[10,11] Kenneth Muir,[12] David Neal,[13] Graham G Giles,[14-16] Edward L Giovannucci,[17-19] Meir Stampfer,[17-19] Brandon L Pierce,[20-22] Joellen M Schildkraut,[23] Shaneda Warren Andersen,[24] Deborah Thompson,[25] Wei Zheng,[26] Peter Kraft[1], Konstantinos K Tsilidis[3,4*] on behalf of the PRACTICAL, CRUK, BPC3, CAPS and PEGASUS consortia

# Xia Jiang and Niki L Dimou share the first authorship.

**Correspondence to:** Konstantinos K Tsilidis, Senior lecturer. Faculty of Medicine, School of Public Health, Praed Street, St Mary's Campus, Imperial College London, UK.

Phone: +30 2651007734; Fax: +30 2651007853; Email: k.tsilidis@imperial.ac.uk

**Word count**: 3310

**Supply Medical Subject Headings**: Vitamin D, circulating concentrations, malignancy, breast cancer, prostate cancer, Mendelian randomization.

**Statements:** The references have been checked for accuracy and completeness. Dr. Xia Jiang will act as guarantor for the paper. We did not cite personal communications in the manuscript. This material has not been published previously in a substantively similar form. The manuscript has been carefully edited by someone with an excellent mastery of the English language.

1. Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health. Boston, USA.
2. Cardiovascular epidemiology unit, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

3. Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, Ioannina, Greece.
4. Department of Epidemiology and Biostatistics, The School of Public Health, Imperial College London, London, UK.
5. Bristol Medical School, Department of Population Health Sciences, University of Bristol, Bristol, UK.
6. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
7. National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and the University of Bristol, Bristol, UK.
8. International Agency for Research on Cancer, Lyon, France.
9. Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK.
10. The Institute of Cancer Research, London, SM2 5NG, UK.
11. Royal Marsden NHS Foundation Trust, London, SW3 6JJ, UK.
12. Institute of Population Health, University of Manchester, Oxford Road, Manchester, M13 9PL, UK.
13. Nuffield Department of Surgery, University of Oxford, Room 6603, Level 6, John Radcliffe Hospital, Headley Way, Headington, Oxford, OX3 9DU, UK and University of Cambridge, Department of Oncology, Box 279, Addenbrooke's Hospital, Hills Road Cambridge CB2 0QQ, UK.
14. Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, Australia.

15. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia.
16. Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia.
17. Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
18. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
19. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
20. Department of Public Health Sciences, The University of Chicago, Chicago, IL, USA.
21. Comprehensive Cancer Center, The University of Chicago, Chicago, IL, USA.
22. Department of Human Genetics, The University of Chicago, Chicago, IL, USA.
23. Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA.
24. Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, 2525 West End Avenue, Nashville 37203, TN, USA.
25. Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge, CB1 8RN, UK.

Members from the PRACTICAL, CRUK, BPC3, CAPS and PEGASUS consortia are provided in the Supplement notes. Information of the PRACTICAL consortium can be found at http://practical.icr.ac.uk/.

## Abstract

**Background:** Observational studies have suggested an association between circulating vitamin D concentrations [25(OH)D] and risk of breast and prostate cancer, which was not supported by a recent Mendelian randomization analysis comprising 15,748 breast and 22,898 prostate cancer cases. Demonstrating causality has proven challenging, and one common limitation of MR studies is insufficient power.

**Methods:** We aim to determine if circulating concentrations of vitamin D are causally associated with the risk of breast and prostate cancer, by using summary level data from the largest-ever genome-wide association studies conducted on vitamin D (N=73,699), breast cancer ($N_{case}$=122,977) and prostate cancer ($N_{case}$=79,148). We constructed a stronger instrument using six common genetic variants (as compared with the previous four variants), and applied several two-sample MR methods.

**Results:** We found no evidence to support a causal association between 25(OH)D and risk of breast cancer [OR per 25nmol/L increase, 1.02 (95%CI: 0.97-1.08), P=0.47], estrogen receptor (ER)+ [1.00 (0.94-1.07), P=0.99] or ER– [1.02 (0.90-1.16), P=0.75] subsets; prostate cancer [1.00 (0.93-1.07), P=0.99] or the advanced subtype [1.02 (0.90-1.16), P=0.72] using the inverse variance weighted method. Sensitivity analyses did not reveal any sign of directional pleiotropy.

**Conclusion:** Despite its almost five-fold augmented sample size and substantially improved statistical power, our MR does not convincingly support a causal effect of circulating 25(OH)D concentrations on breast or prostate cancer risk. However, we can still not exclude a modest or non-linear effect of vitamin D. Future studies may be designed to understand the effect of vitamin D in subpopulations with a profound deficiency.

**Key words:** serum vitamin D concentrations, malignancy, breast, prostate, Mendelian randomization

**Key message**

We did not find a putative causal role of circulating vitamin D concentrations in the risk of breast or prostate cancer.

We still cannot exclude a modest or non-linear effect of vitamin D on malignant disease.

Future work on vitamin D may be focused on cancer mortality or on subpopulations with a profound deficiency.

## Introduction

Vitamin D is a fat-soluble vitamin and steroid pro-hormone that plays an essential role in bone health.[1] Its biologically active form 1,25-hydroxyvitamin D regulates multiple signaling pathways involved in cell proliferation, apoptosis, differentiation and inflammation, and is therefore believed to have an anti-carcinogenic property.[2]

Epidemiologic evidence on the association of circulating vitamin D and risk of breast and prostate cancer, the two most common malignancies in women and men, remains inconclusive. A meta-analysis aggregating data from 24 prospective studies ($N_{case}$=31,867) identified a pooled relative risk of breast cancer for the highest (>31ng/ml) *vs.* lowest (<18ng/ml) blood 25-hydroxyvitamin D [25(OH)D] of 0.92 (95%CI: 0.83-1.02).[3] Another meta-analysis combining data from 19 prospective studies ($N_{case}$=12,824) found a summary relative risk of 1.04 (95% CI: 1.02-1.06) per 10 ng/ml increment in circulating 25(OH)D concentration.[4] Inferring causality from such studies is challenging because it is difficult to exclude reverse causality, confounding or measurement error.

Although the effect of circulating 25(OH)D on cancer risk can be demonstrated by traditional randomized controlled trials (RCT), large-scale RCTs are not currently widely available due to high cost and long duration. In the Women's Health Initiative trial, 36,282 postmenopausal women were randomized to 400 IU vitamin D or placebo, and a hazard ratio of 0.96 (95%CI: 0.85-1.09, P=0.55) for breast cancer was observed after 7 years' follow up.[5] For prostate cancer, only two comprehensive ongoing trials involving both men and women, the "VITAL" launched in 2010 (N=25,871)[6] and the "D-Health" launched in 2014 (N=21,315),[7] will provide an opportunity to clarify the role of vitamin D on male health outcomes upon completion. Yet, both RCTs are likely to be underpowered given the relatively low prostate cancer incidence.

Mendelian randomization (MR) analysis overcomes the limitations of conventional approaches by using genetic variants (single nucleotide polymorphisms, SNPs) as instrumental variables (IVs) for assessing the causal effect of a risk factor (exposure) on an outcome from observational data.[8] A two-sample MR obtain IV-exposure and IV-outcome associations from two different sets of participants. A recent two-sample MR conducted by Dimitrakopoulou *et al.* included summary results from large collaborative networks ($N_{case}$=70,563) to examine the causal role of vitamin D on seven cancers (breast [$N_{case}$=15,748], prostate [$N_{case}$=22,898], lung [$N_{case}$=12,537], colorectal [$N_{case}$=11,488], ovarian [$N_{case}$=4,369], pancreatic [$N_{case}$=1,896] and neuroblastoma [$N_{case}$=1,627]). This study observed little evidence that genetically predicted 25(OH)D was associated with the risk of any cancer [OR per 25nmol/L increase for breast: 1.05 (95%CI: 0.89-1.24); prostate: 0.89 (95%CI: 0.77-1.02)].[9] However, the Dimitrakopoulou *et al.* study used four genetic variants identified from an earlier SUNLIGHT consortium genome-wide association study (GWAS).[10] With the rapid expansion in sample sizes of GWAS, two additional vitamin D associated loci have been recently identified.[11] Incorporating these loci could improve the strength of genetic instrument and both the accuracy and precision of MR estimates. The statistical power can be further improved by using summary genetic data for breast and prostate cancer from recent larger GWASs.[12–14]

Therefore, we conducted an updated two-sample MR analysis to examine the effect of 25(OH)D on breast and prostate cancer. Six genetic variants associated with plasma 25(OH)D concentration were used as IVs. Summary statistics for the IV-exposure were extracted from the largest vitamin D GWAS involving 73,699 individuals.[11] Summary statistics for the IV-outcome were extracted from the largest GWASs for breast (122,977 cases and 105,974 controls) and prostate cancer (79,148 cases and 61,106 controls) conducted by the OncoArray network.[15]

## Methods

### Data for IV-exposure

We retrieved summary data for the association between six SNPs and circulating 25(OH)D concentration from the SUNLIGHT meta-GWAS involving 79,366 discovery samples and 42,757 replication samples of European ancestry. Genome-wide analyses were performed within each cohort according to a uniform analysis plan. Specifically, additive genetic models using linear regression on natural-log transformed 25(OH)D were fitted, and a fixed-effects inverse variance weighted meta-analysis across the contributing cohorts was performed, with control for population structure within each cohort and quality control thresholds of minor allele frequency (MAF)>0.05, imputation info score>0.8, Hardy-Weinberg equilibrium (HWE)>$1\times10^{-6}$, and a minimum of 10,000 individuals contributing to each reported SNP-phenotype association. Information regarding the quality control and statistical analyses has been reported previously.[11]

Among the six SNPs, four were previously identified as being robustly associated with vitamin D (rs3755967 at *GC*, rs12785878 at *NADSYN1/DHCR7*, rs10741657 at *CYP2R1*, rs17216707 at *CYP24A1*). The other two were newly identified by the discovery sample (rs10745742 at *AMDHD1*, rs8018720 at *SEC23A*) and validated in the replication sample.

### Data for IV-outcome

We retrieved summary data for associations of the six vitamin D proxy SNPs with breast and prostate cancer from the currently largest meta-GWAS of these outcomes conducted by the OncoArray network,[15] a large-scale collaboration that was established to understand the genetic architecture and biological mechanisms underlying five common cancers (breast, prostate, ovarian, colorectal and lung cancer). A total of 447,705 individuals of European ancestry were

genotyped on a custom Illumina array, and imputed to the 1000 Genomes Project reference panel. For each cancer type, results from individual GWAS were combined by fixed-effects inverse variance weighted meta-analysis, with control for population stratification within each cohort and quality control thresholds of MAF>0.01, imputation info score>0.3 and HWE>$1\times10^{-12}$. Information regarding the quality control and statistical analyses has been reported previously.[12–14]

For breast cancer, 122,977 cases (105,974 controls) were involved, of which, 69,501 were estrogen receptor (ER)+ cases and 21,468 were ER– cases; for prostate cancer, 79,148 cases (61,106 controls) were involved, of which 15,167 were diagnosed with advanced prostate cancer (defined as Gleason Score 8+ or death from the disease or metastatic disease (i.e. M1) or prostate specific antigen values > 100 ng/ml). Our current sample size was 4.6-fold higher than that in Dimitrakopoulou *et al.*[9].

**Statistical analysis**

MR uses SNPs as proxies for risk factors, assuming that SNPs are randomly allocated at conception, mirroring a randomization process; and that SNPs always precede disease onset thus eliminating reverse causality. Three important assumptions need to be met to ensure a valid IV.[16] The first is the relevance assumption, that the IVs should be ~~strongly~~ associated with the exposure; the second assumption requires no association between IVs and confounders of the exposure-outcome relationship; and the third is the exclusion restriction assumption, indicating that genetic variants should affect the outcome only through the exposure. If all MR assumptions are satisfied then a causal estimation can be made based on the observed IV-exposure and IV-outcome association.

We conducted a two-sample MR to test for a potential causal relationship between circulating 25(OH)D and risk of breast and prostate cancer. We also investigated cancer subtypes including ER+ and ER– breast cancer and advanced prostate cancer. We applied a number of MR methods including an inverse variance weighted average approach (IVW),[17] a maximum likelihood method,[18] MR-Egger regression,[19] and a weighted median approach.[20]

For each of the *k* genetic variants (IVs), the estimate of genetic association with the exposure is represented as $\hat{\beta}_{Xk}$ with standard error $\sigma_{Xk}$, and the estimate of genetic association with the outcome is represented as $\hat{\beta}_{Yk}$ with standard error $\sigma_{Yk}$. The IVW estimator can be motivated as a weighted average of the ratio estimates $\frac{\hat{\beta}_{Yk}}{\hat{\beta}_{Xk}}$ for each variant, weighted using the reciprocal of an approximate expression for their asymptotic variance $\frac{\sigma_{Yk}^2}{\hat{\beta}_{Xk}^2}$, as shown by the formula below. To evaluate potential heterogeneity among causal effects of different variants, we employed the Q test, P-value < 0.05 was considered as the existence of heterogeneity.

$$\text{The causal estimate } \hat{\beta}_{IVW} = \frac{\sum_k \hat{\beta}_{Xk}\hat{\beta}_{Yk}\sigma_{Yk}^{-2}}{\sum_k \hat{\beta}_{Xk}^2\sigma_{Yk}^{-2}}$$

$$\text{The approximate standard error of the estimate } \mathrm{se}\left(\hat{\beta}_{IVW}\right) = \sqrt{\frac{1}{\sum_k \hat{\beta}_{Xk}^2\sigma_{Yk}^{-2}}}$$

Complementary to IVW, we also adopted the maximum likelihood method. When the genetic associations with the exposure are precisely estimated, both approaches give identical results. When there is considerable imprecision in the estimates, causal effect estimates from the IVW are over-precise, whereas the likelihood method gives appropriately-sized confidence intervals.

In addition, we performed MR-Egger regression to test for bias due to directional pleiotropy, where the average of the direct effects of the tested genetic variants on outcome is non-zero (i.e. a violation of exclusion restriction assumption). Under the INstrument Strength Independent of Direct Effect (InSIDE) assumption, the intercept of a weighted regression of $\hat{\beta}_{yk}$ on $\hat{\beta}_{xk}$ will be different than zero in the presence of directional pleiotropy, and the slope of that regression will be a consistent estimate of the causal effect of X on Y.[21] Complementary to MR-Egger, we also employed a weighted median method to derive causal effect estimates. This method provides consistent estimates even when up to 50% of the analyzed genetic variants are invalid IVs. We also employed a multivariable MR approach[22] to adjust for potential horizontal pleiotropy acting in particular through BMI, because rs10741657 has been associated with BMI (P=0.01), and BMI has been associated with 25(OH)D concentrations[23] and breast cancer risk[24]. Publicly available genetic data for BMI were retrieved from the GIANT consortium for 339,000 individuals (95% were of European descent).[25]

Further, we performed a sensitivity analysis where we excluded one SNP at-a-time, and performed IVW on the rest five SNPs to identify the potential influence of outlying variants on the estimates. Finally, we estimated the power of our study according to a method suggested by Brion *et al.*[26]. The six 25(OH)D associated SNPs collectively explained 2.84% of the variance of circulating vitamin D concentration. We fixed the type-I error rate at 0.05.

## Results

Table 1 shows the sample size in the current analysis for each cancer and subtype. The number of overall breast cancer cases was 122,977, of which 69,501 were ER+ cases and 21,468 were ER− cases. The number of overall prostate cancer cases was 79,148, of which 15,167 had

advanced disease. Under the current sample size, our study had 80% power to detect a causal effect of a relative 7% (i.e. ORs of 0.93 or less) decrease in breast cancer risk per 25nmol/L increase of 25(OH)D and 8% for prostate cancer (i.e. ORs of 0.92); corresponding estimates for ER+ breast cancer, ER– breast cancer and advanced prostate cancer were 8%, 12% and 13% relative reductions (i.e. ORs of 0.92, 0.88 and 0.87). We also presented power estimations for a range of proportions of 25(OH)D variation explained by the six genetic variants.

Table 2 presents the information on the association of six SNPs (rs3755967, rs12785878, rs10741657, rs17216707, rs10745742 and rs8018720) with 25(OH) D concentration and cancers. There was no evidence that any of the individual vitamin D associated SNPs were also associated with breast or prostate cancer. We did not find that the genetic instrument for circulating 25(OH)D concentration was associated with the risk of breast [IVW per 25nmol/L increase: 1.02 (95%CI: 0.97-1.08)] or prostate cancer [IVW: 1.00 (95%CI: 0.93-1.07)], breast cancer subtypes [ER+: 1.00 (95%CI: 0.94-1.07); ER–: 1.02 (95%CI: 0.90-1.16)] or advanced prostate cancer [IVW: 1.02 (95%CI: 0.90-1.16)]. The maximum likelihood method generated very similar results. We did not detect heterogeneity among the causal estimates of the six variants ($P_{het}$=0.45 and 0.80 for overall breast and prostate cancer; 0.61 and 0.14 for ER+ and ER– subset; 0.58 for advanced prostate cancer), indicating little evidence for the existence of SNP-specific horizontal pleiotropy. Further, we did not identify aggregated directional pleiotropy using MR-Egger ($P_{intercept}$= 0.92 and 0.88 for overall breast and prostate cancer; 0.82 and 0.70 for ER+ and ER– subset; 0.72 for advanced prostate cancer; the intercepts with 95%CIs are shown in Table 3), although this method has low statistical power when few genetic instruments are used. Consistent with IVW, estimates from MR-Egger and weighed median approach did not provide any evidence of a causal effect of circulating vitamin D on prostate or breast cancer

(Table 3). Multivariable IVW estimates controlling for BMI were almost identical with the classical IVW approach (results not shown).

Similar results were observed in the leave-one-out analysis where we iteratively removed one SNP each time and performed IVW using the remaining five SNPs (Table 4).

## Discussion

In this study, we used an updated and stronger instrumental variable constructed from six SNPs and capitalized on the summary statistics of the largest meta-GWAS conducted for breast and prostate cancers in European populations. We aimed to determine whether the relationship between 25(OH)D and risk of two cancers was causal employing a range of two-sample MR methods. However, none of these analyses suggested a causal relationship between circulating 25(OH)D concentration and breast or prostate cancer risk.

Despite previous retrospective observational studies suggesting an inverse association between higher circulating 25(OH)D concentrations and breast cancer risk, such a relationship has not been firmly supported by evidence from prospective epidemiological studies. Bauer *et al.* meta-analyzed 9 prospective studies ($N_{case}$=5,206) and identified weak evidence of an association between circulating 25(OH)D and risk of postmenopausal [RR per 5ng/mL (approximately 12.5nmol/L) 0.97 (95%CI: 0.93-1.00)], but not premenopausal [0.99 (95%CI: 0.97-1.04)] breast cancer.[27] Similar results were observed in a meta-analysis conducted by Wang *et al.* that aggregated data over 13 prospective studies ($N_{case}$=9,110), where circulating 25(OH)D was inversely associated with postmenopausal breast cancer [$RR_{highest\ vs.\ lowest\ quantile}$ 0.85 (95%CI: 0.75-0.96)], and with similar but imprecise estimates for premenopausal breast cancer [0.84 (95%CI: 0.52-1.35)].[28] When the sample size was further increased to 31,867 individuals and

with a careful inclusion criteria as in the meta-analysis conducted by Kim *et al.*, little evidence of a significant association was reported for either postmenopausal [$RR_{highest\ vs.\ lowest\ category}$ 0.96 (95%CI: 0.85-1.02)] or premenopausal [0.82 (95%CI: 0.48-1.41)] breast cancer.[3] In agreement with these findings, the Women's Health Initiative (WHI) trial of vitamin D in postmenopausal women did not support a role in breast cancer [$HR_{taker\ vs.\ non\text{-}taker}$ 0.96 (95%CI: 0.85-1.09)].[5] However, it has been argued that personal use of vitamin D supplements (outside of the trial) may have obscured the effect. Among the 43% of WHI women who were not taking personal vitamin D at randomization, a reduced risk of breast cancer was observed comparing calcium and vitamin D supplementation to placebo [HR 0.82 (95%CI: 0.70-0.97)].[29] Although our study had 80% power to detect even smaller effect sizes, ranging from 0.89 to 0.95 per 25nmol/L change, we did not find evidence of an association between genetically determined 25(OH)D concentration and risk of breast cancer.

Results of observational studies investigating 25(OH)D with prostate cancer risk are relatively consistent. Xu *et al.* performed a meta-analysis of 17 nested case-control studies ($N_{case}$=11,380) and observed an increased risk of prostate cancer [$RR_{highest\ vs.\ lowest\ category}$ 1.18 (95%CI: 1.07-1.30), I-square=20.8%].[30] A similar effect-estimate was reported by Gao *et al.* with improved precision in a meta-analysis comprising 19 prospective studies ($N_{case}$=12,824) [$RR_{highest\ vs.\ lowest\ category}$ 1.15 (95%CI: 1.06-1.24), I-square=0%].[4] In line with those findings, a large cohort consortium of 10,018 cases and 11,052 controls examined an unweighted polygenic risk score based on four 25(OH)D associated SNPs (rs2282679 at *GC*, rs6013897 at *CYP24A1*, rs10741657 at *CYP2R1*, rs12785878 at *DHCR7*). This analysis found that a greater number of high vitamin D increasing alleles was associated with an increased risk of aggressive prostate cancer,[31] which might merely reflect metabolic events, molecular or immunological alterations relevant to prostate cancer risk.

Our MR found no evidence of an association between genetically predicated 25(OH)D concentration and risk of overall or advanced prostate cancer.

Our previous MR study did not provide strong evidence of a causal link between 25(OH)D and seven cancers. That study might have been underpowered to identify small effects for some cancers or subgroups,[32] as a positive association for ovarian cancer was shown in an independent MR with larger sample size.[33] Insufficient power is a common limitation of MR studies, because the genetic variants usually explain a modest proportion of phenotypic variance. The four previously reported vitamin D associated SNPs could only explain 3.6%-5.2% of the variance of 25(OH)D;[10,34] and the six vitamin D associated SNPs identified by us, could only explain 2.84% out of the 7.5% overall SNP-heritability calculated using the linkage disequilibrium score regression analysis.[11] Although improvement in the proportion of variability explained by IVs was minimal, our overall statistical power was considerably raised, using data from substantially augmented GWASs of breast and prostate cancer. We had 80% power at an alpha level of 0.05 to identify a 7% relative decreased breast cancer risk (i.e. an OR of 0.93) and an 8% relative decreased prostate cancer risk (i.e. an OR of 0.92) per 25nmol/L increase in circulating 25(OH)D. These effect sizes are comparable to the effects observed in meta-analyses of prospective studies for both cancers. However, it is likely that the true causal effect of 25(OH)D is even weaker – if the true causal effect was less than 3%, a magnitude that is probably of limited clinical relevance – we only had a power of 23% for breast cancer, and 15% for prostate cancer, with our current sample size.

MR provides the opportunity to make causal inference between an exposure and an outcome using observational data. The validity of causal estimates requires several assumptions to be satisfied. We selected the most significant independent SNPs identified by the largest 25(OH)D

GWAS, so all were robustly associated with the exposure of interest. The six variants combined constitute a strong instrument, with an *F* statistic of 387. This would minimize any bias due to using a weak instrument in the analysis. Secondly, none of the six instrumental variables (or the proxy SNPs in high linkage disequilibrium with the six IVs at $r^2 \geq 0.8$) used in our analysis were cited by the NHGRI-EBI Catalog of published GWAS as associated with known confounders of cancer risk, such as BMI, smoking, alcohol consumption, mammographic density or inflammation at $\alpha = 10^{-5}$ level. However, due to the lack of individual data, we were not able to test the association of genetic instruments with other confounders such as physical activity, hormonal and lifestyle related factors, which are usually captured by questionnaires. Finally, the major assumption is the exclusion restriction that the genetic determinants affect cancer only through vitamin D concentrations. Violation of this assumption is unlikely as we employed a range of methods known to control for pleiotropy, and obtained highly consistent results. However, we are also aware that MR-Egger regression with only six SNPs could be underpowered to identify pleiotropy. All MR studies so far have only tested the linear effect of circulating vitamin D concentations in the general population. Future studies may be designed to understand the effect of vitamin D in subpopulations with a profound deficiency (non-linear effects), as well as to investigate the causal role of vitamin D in cancer progression or death.

In conclusion, although a very small causal effect of circulating 25(OH)D concentration on breast and prostate cancers cannot be ruled out, our updated analysis, despite its almost five-fold augmented sample size and substantially improved overall statistical power, provides no evidence in support of a causal relationship between circulating concentrations of 25(OH)D and the risk of breast or prostate cancer.

**Funding**

This work was supported by the Veteskapsrådet International Postdoc grant (XJ), the World Cancer Research Fund International Regular Grant Programme (WCRF 2014/1180) (KKT), the World Cancer Research Fund International grant (WCRF 2015/1421) (SL), the National Cancer Institute Grant (R00 CA207848) (SWA), the Cancer Research UK Programme Grant, the Integrative Cancer Epidemiology Programme (C18281/A19169) (RMM), and the National Institute for Health Research (NIHR) Bristol Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

**Conflict of interest** None declared.

## References

1. Doppelt SH. Vitamin D, rickets, and osteomalacia. *Orthop Clin North Am*. 1984 Oct;**15**(4):671–686.
2. Feldman D, Krishnan AV, Swami S, Giovannucci E, Feldman BJ. The role of vitamin D in reducing cancer risk and progression. *Nat Rev Cancer*. 2014 May;**14**(5):342–357.
3. Kim Y, Je Y. Vitamin D intake, blood 25(OH)D levels, and breast cancer risk or mortality: a meta-analysis. *Br J Cancer*. 2014 May 27;**110**(11):2772–2784.
4. Gao J, Wei W, Wang G, Zhou H, Fu Y, Liu N. Circulating vitamin D concentration and risk of prostate cancer: a dose-response meta-analysis of prospective studies. *Ther Clin Risk Manag*. 2018;**14**:95–104.
5. Brunner RL, Wactawski-Wende J, Caan BJ, et al. The Effect of Calcium plus Vitamin D on Risk for Invasive Cancer: Results of the Women's Health Initiative (WHI) Calcium Plus Vitamin D Randomized Clinical Trial. *Nutrition and Cancer*. 2011 Aug 1;**63**(6):827–841.
6. Manson JE, Bassuk SS, Lee I-M, et al. The VITamin D and OmegA-3 TriaL (VITAL): rationale and design of a large randomized controlled trial of vitamin D and marine omega-3 fatty acid supplements for the primary prevention of cancer and cardiovascular disease. *Contemp Clin Trials*. 2012 Jan;**33**(1):159–171.
7. Neale RE, Armstrong BK, Baxter C, et al. The D-Health Trial: A randomized trial of vitamin D for prevention of mortality and cancer. *Contemp Clin Trials*. 2016;**48**:83–90.
8. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003 Feb 1;**32**(1):1–22.
9. Dimitrakopoulou VI, Tsilidis KK, Haycock PC, et al. Circulating vitamin D concentration and risk of seven cancers: Mendelian randomisation study. *BMJ*. 2017 31;**359**:j4761.
10. Wang TJ, Zhang F, Richards JB, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet*. 2010 Jul 17;**376**(9736):180–188.
11. Jiang X, O'Reilly PF, Aschard H, et al. Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. *Nat Commun*. 2018 Jan 17;**9**(1):260.

12. Michailidou K, Lindström S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017 Oct 23;
13. Milne RL, Kuchenbaecker KB, Michailidou K, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet*. 2017 Oct 23;
14. Schumacher F, Al Olama A, Berndt S. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nature Genetics*. **In press**.
15. Amos CI, Dennis J, Wang Z, et al. The OncoArray Consortium: a Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev*. 2016 Oct 3;
16. Zheng J, Baird D, Borges M-C, et al. Recent Developments in Mendelian Randomization Studies. *Curr Epidemiol Rep*. 2017;**4**(4):330–345.
17. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC- InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol*. 2015 Jul;**30**(7):543–552.
18. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013 Nov;**37**(7):658–665.
19. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015 Apr;**44**(2):512–525.
20. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016 May;**40**(4):304–314.
21. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol*. 2017;**32**(5):377–389.
22. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol*. 2015 Feb 15;**181**(4):251–260.
23. Pourshahidi LK. Vitamin D and obesity: current perspectives and future directions. *Proc Nutr Soc*. 2015 May;**74**(2):115–124.
24. Guo Y, Warren Andersen S, Shu X-O, et al. Genetically Predicted Body Mass Index and Breast Cancer Risk: Mendelian Randomization Analyses of Data from 145,000 Women of European Descent. *PLoS Med*. 2016 Aug;**13**(8):e1002105.
25. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015 Feb 12;**518**(7538):197–206.
26. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol*. 2013 Oct 1;**42**(5):1497–1501.
27. Bauer SR, Hankinson SE, Bertone-Johnson ER, Ding EL. Plasma vitamin D levels, menopause, and risk of breast cancer: dose-response meta-analysis of prospective studies. *Medicine (Baltimore)*. 2013 May;**92**(3):123–131.
28. Wang D, Vélez de-la-Paz OI, Zhai J-X, Liu D-W. Serum 25-hydroxyvitamin D and breast cancer risk: a meta-analysis of prospective studies. *Tumour Biol*. 2013 Dec;**34**(6):3509–3517.
29. Bolland MJ, Grey A, Gamble GD, Reid IR. Calcium and vitamin D supplements and health outcomes: a reanalysis of the Women's Health Initiative (WHI) limited-access data set. *Am J Clin Nutr*. 2011 Oct;**94**(4):1144–1149.

30. Xu Y, Shao X, Yao Y, et al. Positive association between circulating 25-hydroxyvitamin D levels and prostate cancer risk: new findings from an updated meta-analysis. *J Cancer Res Clin Oncol*. 2014 Sep;**140**(9):1465–1477.
31. Mondul AM, Shui IM, Yu K, et al. Genetic variation in the vitamin d pathway in relation to risk of prostate cancer--results from the breast and prostate cancer cohort consortium. *Cancer Epidemiol Biomarkers Prev*. 2013 Apr;**22**(4):688–696.
32. Manousaki D, Richards JB. Low vitamin D levels as a risk factor for cancer. *BMJ*. 2017 31;**359**:j4952.
33. Ong J-S, Cuellar-Partida G, Lu Y, et al. Association of vitamin D levels and risk of ovarian cancer: a Mendelian randomization study. *Int J Epidemiol*. 2016;**45**(5):1619–1630.
34. Ahn J, Yu K, Stolzenberg-Solomon R, et al. Genome-wide association study of circulating vitamin D levels. *Hum Mol Genet*. 2010 Jul 1;**19**(13):2739–2745.

**Table1. Number of cancer cases and controls and statistical power in Mendelian randomization study of circulating vitamin D concentration and risk of breast and prostate cancer.**

| Cancer type | Cases | Controls | Total | Proportion of cases | Minimum detectable odds ratio | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $R^2$=0.01 | $R^2$=0.02 | $R^2$=0.03 | $R^2$=0.04 | $R^2$=0.05 |
| **Breast cancer** | | | | | | | | | |
| Overall | 122977 | 105974 | 228951 | 0.54 | 0.89/1.12 | 0.92/1.09 | 0.935/1.069 | 0.943/1.06 | 0.948/1.05 |
| ER-positive | 69501 | 95042 | 164543 | 0.42 | 0.87/1.15 | 0.905/1.104 | 0.922/1.085 | 0.932/1.073 | 0.939/1.065 |
| ER-negative | 21468 | 100594 | 122062 | 0.18 | 0.80/1.25 | 0.855/1.169 | 0.882/1.134 | 0.897/1.11 | 0.908/1.10 |
| **Prostate cancer** | | | | | | | | | |
| Overall | 79148 | 61106 | 140254 | 0.56 | 0.86/1.16 | 0.90/1.11 | 0.92/1.09 | 0.93/1.08 | 0.94/1.07 |
| Advanced | 15167 | 58308 | 73475 | 0.21 | 0.79/1.26 | 0.85/1.18 | 0.87/1.15 | 0.88/1.13 | 0.90/1.11 |

ER: estrogen receptor. Minimum detectable odds ratio per 25nmol/L increase/decrease in 25(OH)D concentration: assume 80% power, 5% alpha level, and that 1% to 5% of 25(OH)D variance is explained by the six SNPs used in the current paper.

**Table2. Characteristics of GWAS-identified genetic variants associated with circulating 25(OH)D concentrations and their effects in the breast and prostate cancer.**

| SNPs | Chr: pos (hg19) | Locus | Effect / Other allele | Circulating Vitamin D Levels | | Overall Breast Cancer | | ER-positive Breast Cancer | | ER-negative Breast Cancer | | Prostate Cancer | | Advanced Prostate Cancer | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Beta (SE) | P-value | Beta (SE) | P-value | Beta (SE) | P-value | Beta (SE) | P-value | Beta (SE) | P-value | Beta (SE) | P-value |
| rs3755967 | 4:72609398 | GC | T/C | -0.089 (0.0023) | 4.74E-343 | -0.0031 (0.0069) | 0.65 | 0.0012 (0.0082) | 0.89 | -0.0103 (0.0125) | 0.41 | -0.0029 (0.0091) | 0.75 | -0.0114 (0.0158) | 0.47 |
| rs10741657 | 11:14914878 | CYP2R1 | A/G | 0.031 (0.0022) | 2.05E-46 | 0.0075 (0.0063) | 0.23 | 0.0015 (0.0075) | 0.84 | 0.012 (0.0115) | 0.30 | 0.0031 (0.0082) | 0.71 | 0.0122 (0.0142) | 0.39 |
| rs12785878 | 11:71167449 | NADSYN1 / DHCR7 | T/G | 0.036 (0.0022) | 3.80E-62 | -0.0016 (0.0070) | 0.82 | 0.0054 (0.0083) | 0.52 | -0.0245 (0.0127) | 0.054 | -0.0111 (0.0090) | 0.22 | -0.0243 (0.0154) | 0.12 |
| rs10745742* | 12:96358529 | AMDHD1 | T/C | 0.0190 (0.0020) | 2.10E-20 | 0.0044 (0.0063) | 0.49 | -0.0011 (0.0075) | 0.88 | 0.0097 (0.0115) | 0.40 | 0.0006 (0.0082) | 0.94 | 0.0000 (0.0142) | 0.99 |
| rs8018720* | 14:39556185 | SEC23A | C/G | -0.019 (0.0027) | 1.11E-11 | 0.0132 (0.0083) | 0.11 | 0.0162 (0.0099) | 0.10 | 0.0222 (0.0152) | 0.14 | -0.0071 (0.0108) | 0.51 | -0.0065 (0.0186) | 0.73 |
| rs17216707 | 20:52732362 | CYP24A1 | T/C | 0.026 (0.0027) | 8.14E-23 | 0.0062 (0.0082) | 0.45 | 0.0063 (0.0098) | 0.53 | 0.0055 (0.0150) | 0.71 | -0.0040 (0.0106) | 0.70 | -0.0026 (0.0182) | 0.88 |
| Reference | | | | Jiang, et al. 2018 | | Michailidou, et al. 2017 | | Michailidou, et al. 2017 | | Milne, et al. 2017 | | Schumacher, et al. 2018 | | Schumacher, et al. 2018 | |

*Novel vitamin D GWAS-identified SNPs, beta-coefficients were extracted from the pooled analysis (discovery dataset + replication dataset). ER: estrogen receptor.
The beta coefficients for the association between SNPs and circulating vitamin D are based on per effect allele per unit change in log 25(OH)D. To enable better comparison with results from observational studies, we run MR analyses after transforming these beta coefficients into the natural scale (nmol/L) using a formula suggested by Rodriguez-Barranco et al. BMC Med Res Methodol 2017 Mar 17;17(1):44.

**Table 3. Mendelian Randomization estimates between genetically predicted 25(OH)D concentrations and cancer risk using multiple 25(OH)D GWAS-identified variants.**

| Cancer type | Method | OR | 95%CI | P-value | $P_{int}$ or $P_{het}$ # |
|---|---|---|---|---|---|
| **Breast** | | | | | |
| Overall | Inverse-variance weighted | 1.02 | (0.97, 1.08) | 0.47 | 0.45 |
| Overall | Maximum likelihood | 1.02 | (0.97, 1.08) | 0.47 | NA |
| Overall | MR-Egger | 1.02 | (0.91, 1.13) | 0.78 | 0.92 |
| | MR-Egger intercept | | | 0.001(–0.010-0.012) | |
| Overall | Weighted Median | 1.02 | (0.96, 1.08) | 0.51 | NA |
| ER-positive | Inverse-variance weighted | 1.00 | (0.94, 1.07) | 0.99 | 0.61 |
| ER-positive | Maximum likelihood | 1.00 | (0.94, 1.07) | 0.99 | NA |
| ER-positive | MR-Egger | 1.01 | (0.90, 1.14) | 0.85 | 0.82 |
| | MR-Egger intercept | | | –0.001(–0.013-0.011) | |
| ER-positive | Weighted Median | 1.00 | (0.93, 1.07) | 0.99 | NA |
| ER-negative | Inverse-variance weighted | 1.02 | (0.90, 1.16) | 0.75 | 0.14 |
| ER-negative | Maximum likelihood | 1.02 | (0.90, 1.16) | 0.75 | NA |
| ER-negative | MR-Egger | 1.06 | (0.83, 1.37) | 0.63 | 0.70 |
| | MR-Egger intercept | | | –0.005(–0.031-0.021) | |
| ER-negative | Weighted Median | 1.05 | (0.94, 1.18) | 0.36 | NA |
| **Prostate** | | | | | |
| Overall | Inverse-variance weighted | 1.00 | (0.93, 1.07) | 0.99 | 0.80 |
| Overall | Maximum likelihood | 1.00 | (0.93, 1.07) | 0.99 | NA |
| Overall | MR-Egger | 1.01 | (0.88, 1.16) | 0.89 | 0.88 |
| | MR-Egger intercept | | | –0.001(–0.015-0.013) | |
| Overall | Weighted Median | 1.01 | (0.94, 1.10) | 0.73 | NA |
| Advanced | Inverse-variance weighted | 1.02 | (0.90, 1.16) | 0.72 | 0.58 |
| Advanced | Maximum likelihood | 1.02 | (0.90, 1.16) | 0.72 | NA |
| Advanced | MR-Egger | 1.06 | (0.78, 1.43) | 0.62 | 0.72 |
| | MR-Egger intercept | | | –0.004(–0.035-0.026) | |
| Advanced | Weighted Median | 1.05 | (0.91, 1.21) | 0.49 | NA |

NA: not applicable. The odds ratios represent increase/decrease of risk per 25nmol/L increase in 25(OH)D. # $P_{het}$: P-values of Chi-square Q test for heterogeneity were shown; $P_{int}$ P-values of MR-Egger regression test on the intercept.

**Table 4. Mendelian Randomization estimates between genetically predicted 25(OH)D and cancer risk, a sensitivity analysis leaving**

**one SNP out at a time.**

| SNP (left out) | OR | 95%CI | P-value |
|---|---|---|---|
| **Overall Breast Cancer** | | | |
| rs3755967 | 1.03 | (0.93, 1.15) | 0.55 |
| rs10741657 | 1.01 | (0.95, 1.07) | 0.71 |
| rs12785878 | 1.03 | (0.96, 1.09) | 0.42 |
| rs10745742 | 1.02 | (0.96, 1.08) | 0.56 |
| rs8018720 | 1.03 | (0.97, 1.09) | 0.33 |
| rs17216707 | 1.02 | (0.96, 1.08) | 0.57 |
| **ER-positive Breast Cancer** | | | |
| rs3755967 | 1.02 | (0.90, 1.14) | 0.80 |
| rs10741657 | 1.00 | (0.93, 1.07) | 0.96 |
| rs12785878 | 0.99 | (0.93, 1.06) | 0.84 |
| rs10745742 | 1.00 | (0.94, 1.07) | 0.96 |
| rs8018720 | 1.01 | (0.94, 1.08) | 0.80 |
| rs17216707 | 1.00 | (0.93, 1.06) | 0.91 |
| **ER-negative Breast Cancer** | | | |
| rs3755967 | 0.95 | (0.74, 1.22) | 0.71 |
| rs10741657 | 1.00 | (0.87, 1.16) | 0.95 |
| rs12785878 | 1.06 | (0.95, 1.18) | 0.27 |
| rs10745742 | 1.01 | (0.88, 1.17) | 0.85 |
| rs8018720 | 1.03 | (0.91, 1.17) | 0.61 |
| rs17216707 | 1.02 | (0.88, 1.18) | 0.82 |
| **Overall Prostate Cancer** | | | |
| rs3755967 | 0.97 | (0.85, 1.10) | 0.64 |
| rs10741657 | 1.00 | (0.92, 1.07) | 0.91 |
| rs12785878 | 1.02 | (0.94, 1.10) | 0.66 |
| rs10745742 | 1.00 | (0.93, 1.08) | 0.99 |
| rs8018720 | 1.00 | (0.93, 1.07) | 0.93 |
| rs17216707 | 1.00 | (0.93, 1.08) | 0.93 |
| **Advanced Prostate Cancer** | | | |
| rs3755967 | 0.95 | (0.76, 1.19) | 0.67 |
| rs10741657 | 1.01 | (0.88, 1.15) | 0.92 |
| rs12785878 | 1.07 | (0.93, 1.22) | 0.35 |
| rs10745742 | 1.02 | (0.90, 1.16) | 0.71 |
| rs8018720 | 1.02 | (0.90, 1.16) | 0.76 |
| rs17216707 | 1.03 | (0.90, 1.17) | 0.69 |

The odds ratios represent increase/decrease of risk per 25nmol/L increase in 25(OH)D. Odds ratios calculated using inverse variance weighted method.